Acta Biochimica Polonica

Vol. 44 No. 1/1997

147–152

QUARTERLY

Short communication

Identification of a microsatellite region composed of a long homopurine/homopyrimidine tract surrounded by AT-rich sequences upstream of the rat stress-inducible hsp70.1 gene*

Katarzyna Lisowska✉, Tomasz Loch, Anna Fiszer-Kierzkowska, Dorota Ścieglińska and Zdzisław Krawczyk

Department of Tumor Biology, Institute of Oncology, 44-100 Gliwice, Poland

Received: 20 February, 1996; revised: 7 January, 1997; accpeted: 10 January, 1997

Key words: heat shock, hsp70.1, microsatellite sequence, H-DNA, cruciform, homopurine/homopyrimidine

A DNA region containing several repetitive motifs has been detected about 1.9 kbp upstream of the transcription unit of the rat stress-inducible hsp70.1 gene. The most interesting element of this area is a microsatellite sequence $(GA)_6CAG(TC)_{24}$ that consists of an inverted repeat partially overlapping with the long homopurine/homopyrimidine tract (Pu/Py). DNA molecule within the described sequence can theoretically adopt alternate, non-B structures (H-DNA or cruciform) containing single-stranded regions. This microsatellite region is flanked by AT-rich sequences containing several poly(A) tracts. The longest of them with a possible potential to destabilize a double-stranded DNA helix is localized around 160 bp downstream the $(GA)_6CAG(TC)_{24}$. The DNA fragment containing sequences described above was subcloned into the pUC19 vector and the resulting plasmid was subjected to the standard S1 susceptibility assay. Preliminary mapping of the S1 cleavage site indicates for the formation of the non-B-DNA structure within the Pu/Py tract. This is to our knowledge a first report on the existence of a complex microsatellite region upstream the 5′-end of the hsp70 gene in mammals.

The homopurine/homopyrimidine DNA stretches are commonly found in mammalian genomes within 5′ and 3′ end flanking sequences as well as within the coding regions of various genes [1]. Such DNA regions composed of Pu/Py repetitive sequences are suggested to form triple-stranded DNA structure called H-DNA [2, 3]. In the human genome the frequency of finding sequences with a potential to form H-DNA is about 1 in

*This work has been supported by a grant from the Polish National Cancer Program CPBR 11.5, by the State Committee for Scientific Research (grant PB 0329/P2/93/05) to Z.K. and by the Foundation for Polish Science (grant BIMOL 76/9) to Z.K.

✉To whom correspondence should be addressed: Katarzyna Lisowska, Department of Tumor Biology, Center of Oncology, Maria Skłodowska-Curie Memorial Institute, Wybrzeże Armii Krajowej 15, 44-100 Gliwice, Poland, tel: 31 10 61 ext. 892, fax: 48+32+313512

Abbreviations: H-DNA, "hinged DNA", triple stranded DNA structure; hsp, heat shock protein; MAR, nuclear matrix attachement sites; MHC, major histocompatibility complex; TBE buffer, Tris/borate/EDTA electrophoresis.

50 kbp [4]. Long Pu/Py tracts once regarded as "junk" DNA are currently considered to be important structures of several possible functions *in vivo*. Such sequences can function either as a negative or positive "*cis*" regulatory element in transcription ([5, 6] and references therein). Another proposed function for the long homopurine/homopyrimidine region is that of the recombination "hot-spot" ([7] and references therein). The Pu/Py regions with a potential to form H-DNA structure are also frequently found within nuclear matrix attachments sites (MAR) ([6, 8] and references therein). Palindromic sequences and inverted repeats can form cruciform structures, which may be involved in origin of replication of DNA ([8] and references therein). Repetitive microsatellite sequences are also suggested to be responsible for increased recombination frequency.

Recently we isolated from a rat genomic library [9] the heat inducible hsp70.1 gene [10, 11]. This gene is one of the three hsp70 genes localized within the rat MHC complex [12]. Our earlier results indicated that the DNA region flanking the 5′ end of the hsp70.1 gene contains repetitive sequences [9]. Here we report preliminary characterization of the long homopolypurine/homopolypyrimidine tract found upstream the transcription initiation site of the hsp70.1 gene.

MATERIALS AND METHODS

DNA sequencing and analysis. Sequencing was performed with the polymerase T7 Kit (Pharmacia) as described earlier [11]. DNA sequence was analyzed by PC/GENE (IntelliGenetics) software, programs DNA-HELIX, RICHREG and PHYSCHEM.

S1 nuclease assay. Plasmid DNA was incubated in the S1 buffer (final concentration: 30 mM sodium acetate, pH 4.5, 50 mM NaCl, 1 mM zinc chloride, and 5% glycerol) for 30 min on ice [13]. S1 nuclease (Pharmacia) was diluted in the buffer containing 10 mM sodium acetate, pH 5.0, 1 mM dithiothreitol, 0.1 mM zinc chloride and 0.001% Triton X-100 [14]. Appropriate concentrations of the nuclease were added to the DNA samples (usually: 5, 10, 25, 50, 100 and 500 u of S1 nuclease were used) and reactions were carried out on ice for 10 min according to [13, 14]. Reactions were stopped by the addition of Tris base and EDTA to the final concentrations of 90 mM and 45 mM, respectively [14].

Mapping of the S1 cleavage site(s). Preliminary mapping of the S1 cleavage site was done by the digestion of the S1 treated plasmids with the *Pst*I enzyme recognizing unique restriction site within both pUC19 and p68/1.0 plasmids. S1 digested plasmids were phenol extracted and ethanol precipitated. DNA was then resuspended in 1 × H buffer (Boehringer) and incubated with the *Pst*I (Boehringer) and with *Hpa*I (Pharmacia) restriction enzymes at 37°C. After incubation the DNA samples were fractionated in 1% agarose gel at 6.5 V/cm in 0.5 × TBE buffer.

RESULTS AND DISCUSSION

We have previously shown that the 4.1 kbp DNA fragment isolated by us from the rat genomic library contained entire transcription unit of the heat inducible hsp70.1 gene [9, 11]. After establishing the nucleotide sequence of the gene [11] we sequenced the upstream DNA region in search for repetitive DNA sequences and possible distal regulatory elements. At the distance of 1.9 kbp upstream the transcription start site of the hsp70.1 gene we found the region of approx. 500 bp in which repetitive sequences of several types were localized (Fig. 1). A central part of this region contains multiple microsatellite sequences. The longest $(GA)_6CAG(TC)_{24}$ sequence is composed of 12 bp long inverted repeat partially overlapping with the 48 bp long homopurine/homopyrimidine tract. This sequence is directly flanked at 5′ and 3′ end by 11 and 9 repeats of TG dinucleotide respectively. Upstream and downstream of the above described microsatellite region several tracts enriched in AT (over 80%) are localized. These AT-rich regions contain multiple poly(A) tracts with the longest of approximately $(A)_{70}$ localized 170 bp downstream the $(GA)_6CAG(TC)_{24}$ sequence (Fig. 1).

In the conditions of superhelical strain long Pu/Py tracts were found to form H-DNA triplex while inverted repeats are able to adopt cruciform structure ([2–4] and references therein). In each of these structures the existence of single stranded regions prone to digestion with S1 nuclease is anticipated. Also the enrichment in AT bases decreases the melting temperature of the double DNA helix and can destabilize the structure of DNA.

above was subcloned in the pUC19 vector (Fig. 2A). Then, the resulting p68/1.0 plasmid as well as the control pUC19 plasmid were digested with increasing concentration of the S1 nuclease and subsequently cleaved with the *Pst*I restriction enzyme (Fig. 2B). The p68/1.0 plasmid was expected to have three potential sites sensitive to nuclease S1. Two of them: the $(GA)_6CAG(TC)_{24}$ region and the region containing long poly(A) tract, are both localized within an insert. Third region —

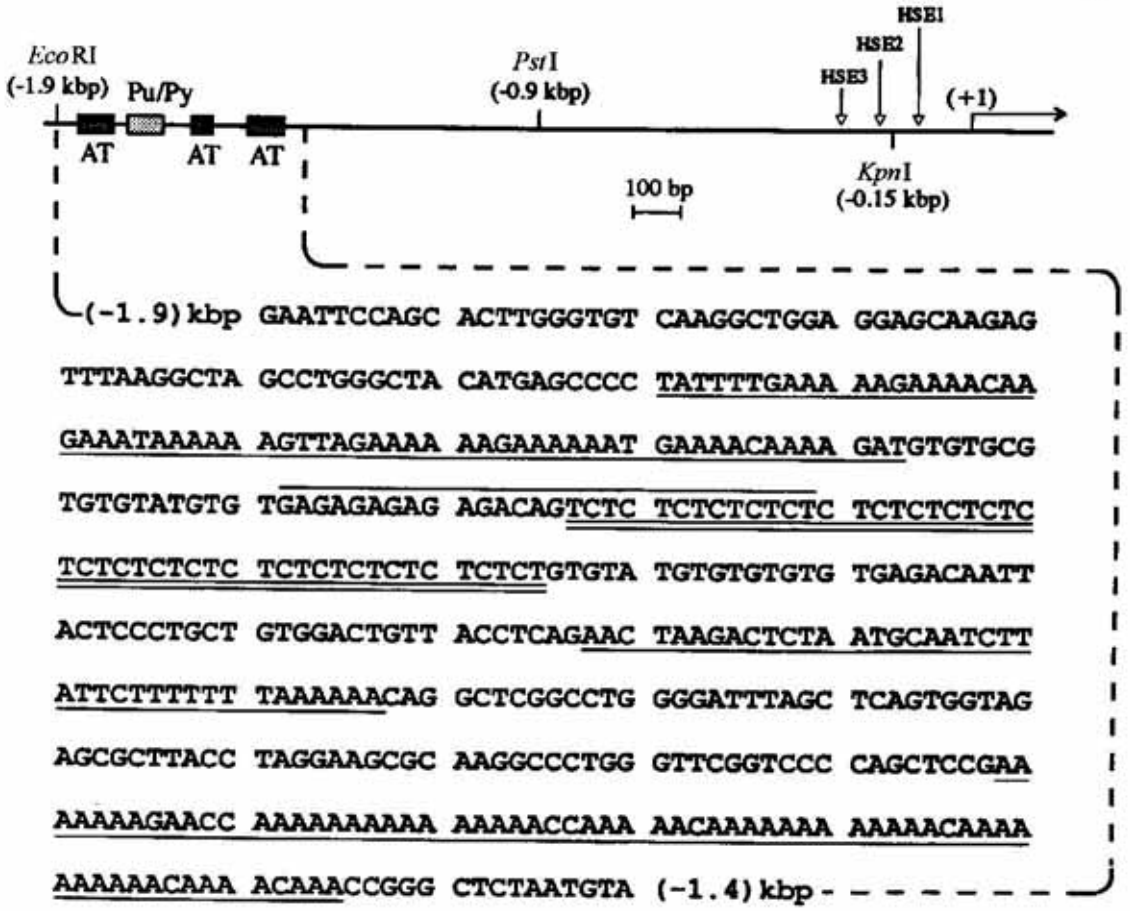

Figure 1. Schematic representation of the far 5′ upstream region of the hsp70.1 gene with part of its nucleotide sequence and approximate position of some restriction sites.

Dashed boxes indicate AT-rich regions and Pu/Py tract. (+1) indicates transcription initiation site and the arrow shows direction of the mRNA synthesis. HSE1, HSE2 and HSE3 indicate so called heat shock elements, regulatory sequences typical for heat shock genes promoter region. DNA sequences enriched in A+T base pairs are underlined. The polypurine/polypyrimidine (Pu/Py) sequence which can potentially adopt triplex conformation is double underlined. Inverted repeat potentially capable of forming cruciform structure is marked by upper line. The sequence has been submitted to EMBL data base with the accession number X74271.

In order to reveal a possibility of non-B-DNA structure formation within the $(AT)..(GA)_6CAG(TC)_{24}..(AT)$ region described above we performed a standard nuclease S1 susceptibility assay [2, 3, 15, 16]. The 1.0 kbp *Eco*RI–*Pst*I restriction fragment containing all repetitive sequences described

composed of three inverted repeats with a potential to form cruciform structures — is localized within the pUC19 vector ([17]; see also Fig. 2A).

It can be seen that the digestion of the p68/1.0 plasmid with S1 nuclease followed by *Pst*I restriction generates the DNA fragment

of approx. 0.85 kbp (Fig. 2B, left panel). In the same conditions the pUC19 vector was only linearized (Fig. 2B, right panel). These data indicate that single stranded region(s) susceptible to nuclease S1 is formed by sequences of the insert and not by inverted repeats region within the pUC19 vector. The size of the DNA restriction fragment cleaved from the p68/1.0 plasmid indicates also that nuclease S1 digests the insert most probably within the Pu/Py stretch and not within the fragment was secondary digested with *Hpa*I enzyme. *Hpa*I cuts once within the insert and there is no restriction site recognized by this enzyme within the vector. Figure 2C shows that *Hpa*I cleaves the S1–*Pst*I fragment producing two bands, from which it may be deduced unambiguously that S1 site lies within the insert and most probably within Pu/Py region. More precise localization of the single-stranded region needs cloning and sequencing of the DNA fragments generated by

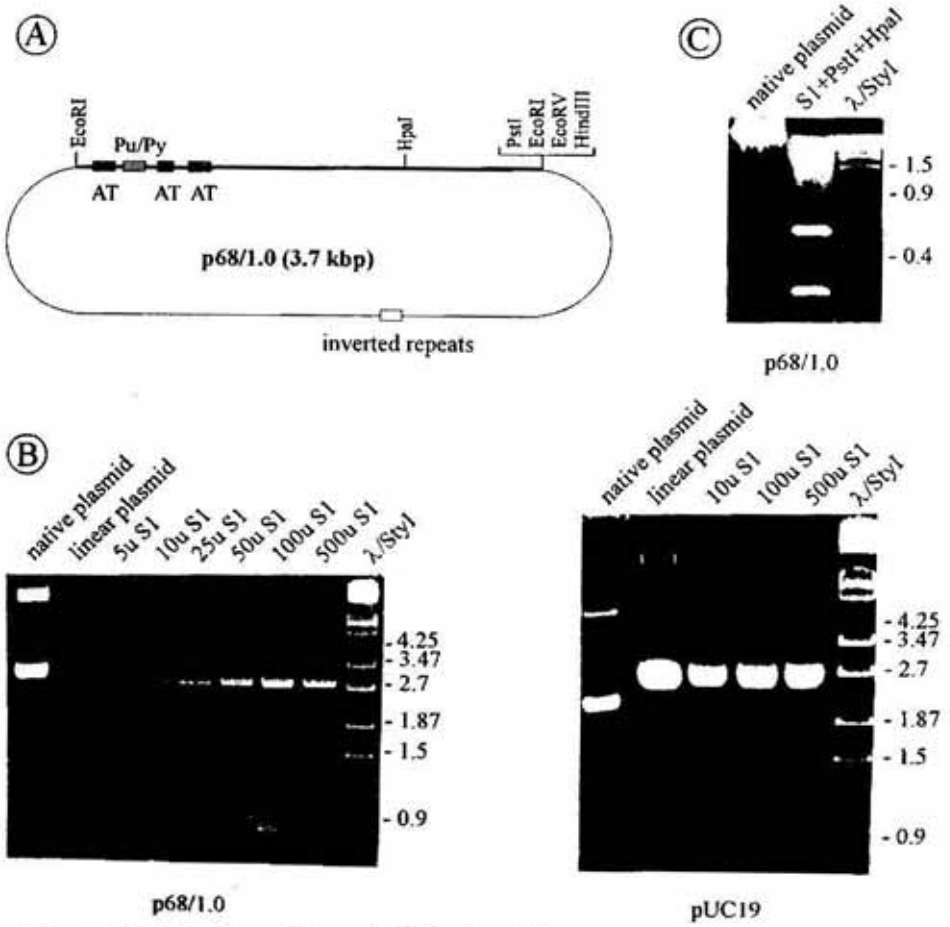

Figure 2. Structure and digestion of the p68/1.0 plasmid.

(A) Structure of the p68/1.0 plasmid. *Eco*RI–PstI restriction fragment of the rat DNA is represented by thick line. pUC19 Vector sequence is represented by thin line. Pu/Py region and AT-rich stretches are shown schematically as dashed boxes. Open box shows the localization of three inverted repeats in the vector sequence (close to one another). These elements are supposed to have a potential to form cruciforms [17]. *Hpa*I is a unique restriction site present in the insert sequence. (B) Double digestion of the p68/1.0 plasmid (left) and pUC19 vector (right) with the S1 nuclease followed by *Pst*I restriction enzyme. (C) Triple digestion of the p68/1.0 plasmid: after S1 and *Pst*I digestion the resulting 0.85 kbp fragment was restricted with *Hpa*I enzyme. Marked lane contains DNA digested with *Sty*I enzyme; size of the marker DNA fragments is given in kbp.

AT-rich region localized downstream this sequence. To further confirm that the above 0.85 kbp fragment is the result of S1 cleavage within Pu/Py region of the insert and not within inverted repeats of the vector, this S1 nuclease. It would also allow us to discuss the structure adopted by the DNA in the microsatellite region.

The role (if any) of the microsatellite region detected by us upstream the rat hsp70.1 gene

is not yet known. As it has been mentioned in the Introduction such sequences with a potential to form either H-DNA or cruciform are thought to be involved in the regulation of gene expression, replication of DNA and recombination.

An intriguing possibility is that the microsatellite region detected by us can be a recombination "hot-spot". In human, mouse and rats two heat inducible hsp70 genes and the spermatid-specific hsc70t gene are localized within the class III MHC region and in these species the MHC-linked hsp70 genes cluster has the same organization. It has been shown that in the mouse genome most recombinations in this part of the MHC complex take place between the hsp70.1 gene and the G7 gene [18]. The DNA sequence in the above mentioned recombination spot in mouse genome is not known, thus it can be only speculated that it can have a similar structure as Pu/Py region found by us upstream of the rat hsp70.1 gene.

The Pu/Py regions with a potential to form H-DNA structure are regarded as possible nuclear matrix attachment sites ([6, 8] and references therein). Indeed, the described DNA region containing the $(GA)_6CAG(CT)_{24}$ sequence surrounded by AT-rich regions strongly resembles the AT-rich type of MARs. Our preliminary data show that the DNA fragment containing repetitive sequences described in this paper binds efficiently nuclear matrix proteins *in vitro* (P. Widłak, K. Lisowska & Z. Krawczyk, unpublished).

Finally, it has to be noted that microsatellite sequences may have a potential application in diagnosis of genomic instability [19–21]. Further characterization of the DNA structure described in this report and determination of its possible *in vivo* function(s) are in progress.

We thank Mrs K. Chorąży and Mrs K. Klyszcz for skillful technical assistance and Mrs H. Waniek for making prints. We thank also Dr M. Rusin for help in verifying the part of the presented nucleotide sequence.

REFERENCES

1. Lu, G. & Ferl, R.J. (1993) *Int. J. Biochem.* **25,** 1529–1537.
2. Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A. & Frank-Kamenetskii, M.D. (1987) *Nature* **330,** 495–497.
3. Htun, H. & Dahlberg, J.E. (1988) *Science* **241,** 1791–1796.
4. Schroth, G.P. & Shing Ho, P.S. (1995) *Nucleic Acids Res.* **23,** 1977–1988.
5. Holde, K. & Zlatanova, J. (1994) *BioEssays* **16,** 59–67.
6. Michel, D., Chatelain, G., Herault, Y., Harper, F. & Brun, G. (1993) *Cell. Mol. Biol. Res.* **39,** 131–140.
7. Santra, M., Danielson, K.G. & Iozzo, R.V. (1994) *J. Biol. Chem.* **269,** 579–587.
8. Boulikas, T. & Kong, C.F. (1993) *J. Cell. Biochem.* **53,** 1–12.
9. Wiśniewski, J. & Krawczyk, Z. (1988) *Mol. Biol. Rep.* **13,** 21–28.
10. Lisowska, K., Wiśniewski, J. & Krawczyk, Z. (1990) *Acta Biochim. Polon.* **37,** 55–58.
11. Lisowska, K., Krawczyk, Z., Widłak, W., Wolniczek, P. & Wiśniewski, J. (1994) *Biochim. Biophys. Acta* **1219,** 64–72.
12. Walter, L., Rauh, F. & Günter, E. (1994) *Immunogenetics* **40,** 325–330.
13. Pahwa, G.S. & Maher, L.J. (1994) *Nucleic Acids Res.* **22,** 1138–1146.
14. Bacolla, A. & Wu, F.Y.-H. (1991) *Nucleic Acids Res.* **19,** 1639–1647.
15. Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V. & Vassart, G. (1985) *Nucleic Acids Res.* **13,** 5127–5137.
16. Mace, H.A.F., Pelham, H.R.B. & Travers, A.A. (1983) *Nature* **340,** 555–557.
17. Lilley, D.M.J. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77,** 6468–6472.
18. Yoshino, M., Sagai, T., Lindahl, K.F., Toyoda, Y., Shirayoshi, Y., Matsumoto, K., Sugaya, K.,

Ikemura, T., Moriwaka, K. & Shiroishi, T. (1994) *Immunogenetics* **40,** 280–286.

19. Leach, D.R.F. (1994) *BioEssays* **16,** 893-899.

20. Loeb, L.A. (1994) *Cancer Res.* **54,** 5059–5063.

21. Rüschoff, J., Bocker, T., Schlegel, J., Stumm, G. & Hofstaedter, F. (1995) *Virchows Archiv* **426,** 215–222.